GIBBS SAMPLING, METROPOLIS HASTINGS & RELATED PROBLEMS

Anaranya Basu
Instructor: Prof.Manisha Pal

November 2023

The following practical problems were given during the lectures of Bayesian Inference by Prof. Dr.Manisha Pal during the third semester class.

1 GIBBS SAMPLING

1.1 Question 1

Suppose the joint distribution of Y and θ is given by-

$$P(Y,\theta) = \binom{n}{Y}\theta^{(Y+a-1)}(1-\theta)^{(n-Y+b-1)}, \quad Y = 0,1,2,\ldots,n, \quad 0 < \theta < 1$$

Use Gibbs sampling to find the marginal distribution of Y given n = 16, a = 2 and b = 4. The initial value of θ can be chosen from a U (0,1) distribution.

Solution:

Given,

$$f_Y(k;n,\theta) = \binom{n}{k} \cdot \theta^k \cdot (1-\theta)^{n-k}$$

and, $\pi(\theta) \propto \theta^{x+a-1}(1-\theta)^{n-x+b-1}$
The Joint density of (Y,θ) is -

$$f(Y,\theta)\pi(\theta) = \binom{n}{Y}\theta^{x+a-1}(1-\theta)^{n-x+a-1}\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}$$

The Marginal of Y is-

$$f(Y) = \binom{n}{Y}\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\frac{\Gamma(a+Y)\Gamma(b+n-Y)}{\Gamma(a+b+n)}$$

The Gibbs Sampling Algorithm is described as follows-
Suppose we have a joint probability distribution $P(X_1, X_2, \ldots, X_n)$ over n variables $X_1, X_2, \ldots, X_n$.

Initialization: Start with initial values for each variable: $(x_1^{(0)}, x_2^{(0)}, \ldots, x_n^{(0)})$.

Iterative Sampling: For $t = 1$ to some predefined number of iterations or until convergence: For each variable X_i:

- Sample $x_i^{(t)}$ from its conditional distribution given the current values of the other variables:

$$x_i^{(t)} \sim P(X_i \mid x_1^{(t)}, x_2^{(t)}, \ldots, x_{i-1}^{(t)}, x_{i+1}^{(t-1)}, \ldots, x_n^{(t-1)})$$

- Update the value of X_i to $x_i^{(t)}$.

Convergence Criteria: Monitor convergence using a suitable criterion, such as assessing the change in sampled values between iterations or using statistical tests.

Termination: Stop the iterations either when convergence is achieved or after a predetermined number of iterations.(Here we have set number of iteration to 1000)

Step 1 :

Now as the conditional distribution of $Y|\theta$ ***is-***

$$f(Y|\theta) = \frac{\binom{n}{Y}\theta^{Y+a-1}(1-\theta)^{n-Y+b-1}}{\sum_{k=0}^{n}\binom{n}{k}\theta^{k+a-1}(1-\theta)^{n-k+b-1}}$$

Step 2:

Now we will Set the initial values of Y and θ. In this case, we can choose the initial value of θ from a uniform distribution between 0 and 1 and the initial value of Y from a binomial distribution with size n and probability θ.

Step 3:

In Next step- For each iteration of the Gibbs sampler, do the following:

- Sample Y from the conditional distribution of Y given θ.
- Sample θ from the conditional distribution of θ given Y.

In the final step we will Calculate the marginal distribution of Y.

After the Gibbs sampler has run for a sufficiently large number of iterations, the marginal distribution of Y can be approximated by the empirical distribution of the sampled Y values.

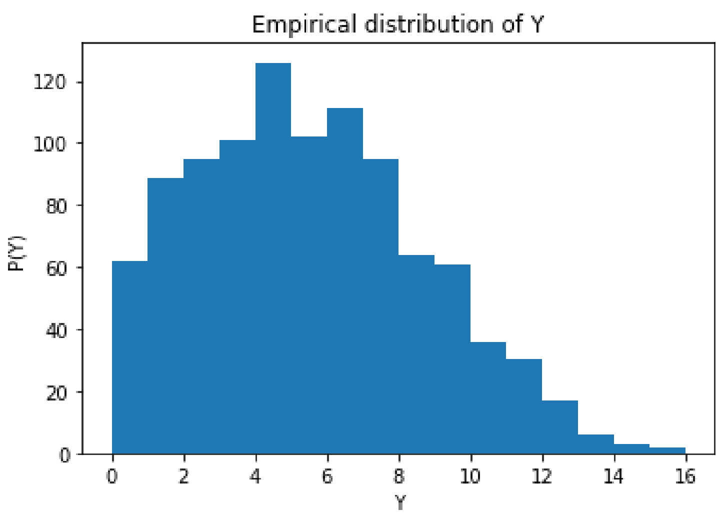
Using R command we get the Gibbs sampler values as follows-

```
   0    1    2    3    4    5    6    7    8    9   10   11   12   13   14   15   16
  34   54   84  119  111   90  114   95   92   69   53   33   20   16    9    4    2
```

And the marginal distribution of Y, which is the probability of each value of Y occurring is-

y	Probability
0	0.020580
1	0.082564
2	0.203356
3	0.274624
4	0.224809
5	0.125616
6	0.053642
7	0.010517
8	0.001329
9	0.000284
10	0.000078
11	0.000017
12	0.000004
13	0.000001
14	0.000000
15	0.000000
16	0.000000

The emperical distribution plot of the marginal of Y is as follows-



The distribution appears to be positively skewed, with a higher concentration of values on the lower end.

The R code is -

```
n <- 16
a <- 2
b <- 4
# Set the number of iterations
M <- 1000
# Initialize the vectors for storing the samples
y_samples <- rep(NA, M)
theta_samples <- rep(NA, M)
# Initialize the starting values
theta <- runif(1)
y <- rbinom(1, size = n, prob = theta)
# Perform Gibbs sampling iterations
for (i in 2:M) {
  # Sample Y from the conditional distribution P(Y | theta)
  y <- rbinom(1, size = n, prob = theta)

  # Sample theta from the conditional distribution P(theta | Y)
  theta <- rbeta(1, shape1 = a + y, shape2 = b + n - y)

  # Store the samples
  y_samples[i] <- y
  theta_samples[i] <- theta
}

# Calculate the marginal distribution of Y
y_counts <- table(y_samples)
y_probs <- y_counts / M

# Print the marginal distribution of Y
print(y_probs)
```

1.2 Question 2

Consider a single observation (y_1, y_2) = (0,0) from a bivariate normally distributed population (θ_1,θ_2, 1, 1, 0.8). Assume the prior distribution of θ_q and θ_2 to be uniform. Apply Gibb's sampler to generate samples on θ_1 and θ_2 from their posterior distributions. Compute their posterior means and variances and show their posterior distributions.

Solution:

The conditional distributions are given by:

$$\pi(\theta_1|\theta_2, y) \propto N(\theta_1; \mu_1, \sigma_1^2)$$

$$\pi(\theta_2|\theta_1, y) \propto N(\theta_2; \mu_2, \sigma_2^2)$$

Where,

- $\mu_1 == E[\theta_1|\theta_2, y]$
- $\sigma_1^2 = Var[[\theta_1|\theta_2, y]$
- $\mu_2 = E[\theta_2|\theta_1, y]$
- $\sigma_2^2 = Var[\theta_2|\theta_1, y]$

Now Let's denote the observation as y = (y_1, y_2) and the parameters as, $\theta = (\theta_1, \theta_2)$. The joint distribution of y and θ is given by:

$$\pi(y, \theta) = f(y|\theta) * \pi(\theta)$$

where $\pi(y, \theta)$ is the bivariate normal likelihood function with parameters $(\theta_1, \theta_2, 1, 10.8)$ and $\pi(\theta)$ is the uniform prior distribution.
The posterior distribution of θ given y is proportional to the product of the likelihood function and the prior distribution:

$$\pi(\theta, y) \propto f(y|\theta) * \pi(\theta)$$

To generate samples from the posterior distribution using Gibbs sampling, we iterate the following steps:

- Sample θ_1 from the conditional distribution $\pi(\theta_1|\theta_2, y)$
  This conditional distribution is proportional to the product of the likelihood function and the prior distribution, with θ_2 fixed, i.e.,

  $$\pi(\theta_1|\theta_2, y) \propto \pi(y_1|\theta_1, \theta_2)\pi(\theta_1)$$

  Since the prior distribution is uniform, the conditional distribution is proportional to the likelihood function:

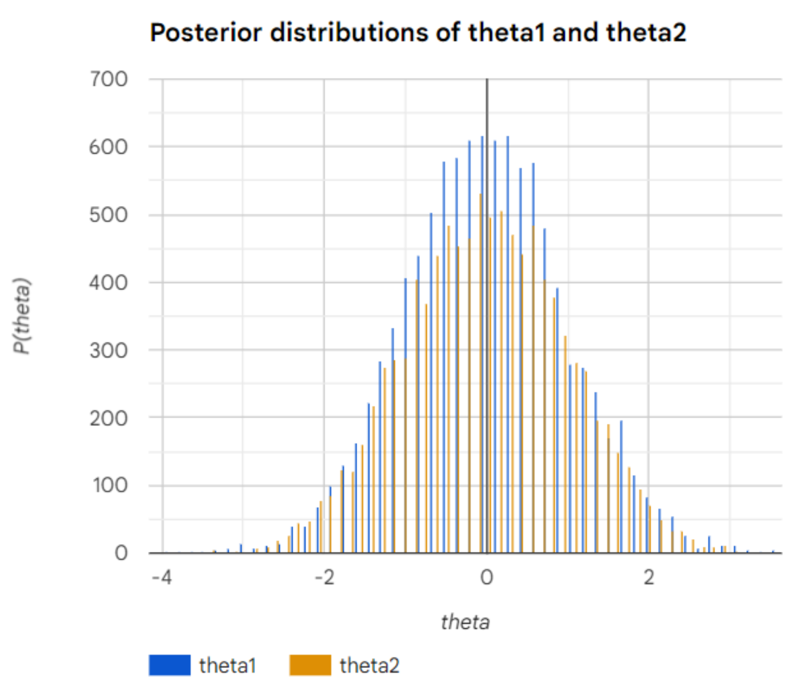
  $$\pi(\theta_1|\theta_2, y) \propto \pi(y_1|\theta_1, \theta_2)$$

- Sample θ_2 from the conditional distribution $\pi(\theta_2|\theta_1, y)$:
  On a similar process, since the prior distribution is uniform, the conditional distribution is proportional to the likelihood function.

Repeat these steps for a sufficient number of iterations to generate samples from the posterior distribution.
By computation we get the values as follows-

```
Posterior mean of theta1: -0.011325169682256702
Posterior variance of theta1: 0.9754217984480076
Posterior mean of theta2: 0.02240552927307522
Posterior variance of theta2: 0.9789807544271096
```

The distribution of the posterior will be visualized from the below histogram:



2 Hierarchical Bayes & Metropolis Hastings:

Let Y_{ij} be coagulation time in seconds for blood drawn from j^{th} animal in the i^{th} group. Altogether 24 animals were randomly allocated to 4 different groups and were fed 4 different diets.

Diet	**Coagulation time**
A	62, 60, 63, 59
B	63, 67, 71, 64, 65, 66
C	68, 66, 71, 67, 67, 68, 68
D	56, 62, 60, 61, 63, 64, 63, 59

Table 1: Coagulation times for different diets

Consider the following linear model, $y_{ij} \sim N(\theta_i, \sigma^2)$, i=1,2,3,4 and j=1(1)$n_i$.Suppose θ_i is unknown for all i,while $\sigma^2 = 16$; Assume prior distribution $\theta_i \sim N(62, 4)$

- Write down the joint Posterior density of all parameters $\theta_1, \theta_2, \theta_3, \theta_4$
- Generate 20 sample observation from the joint posterior density using Metropolis Hastings Algorithm

Solution:

First, the joint posterior density of all parameters $\theta_1, \theta_2, \theta_3, \theta_4$ given the data and the prior information can be found using Bayesian approach:

$$\pi(\theta_1, \theta_2, \theta_3, \theta_4|data) \propto \pi(data|\theta_1, \theta_2, \theta_3, \theta_4)\pi(\theta_1, \theta_2, \theta_3, \theta_4)$$

Given that, $Y_{ij} \sim N(\theta_i, \sigma^2)$ and $\sigma^2 = 16$ and assuming independent priors for each θ_i with $\theta_i \sim N(62, 4)$ the likelihood function can be formulated as the product of normal densities for each group:

$$\pi(\theta_i|data) \propto \prod_{i=1}^{4} \left(\prod_{j=1}^{n_i} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(Y_{ij}-\theta_i)^2}{2\sigma^2}\right)\right) \times \frac{1}{\sqrt{2\pi \times 4}} \exp\left(-\frac{(\theta_i - 62)^2}{2\times 4}\right)$$

Where: Y_{ij} is the coagulation time for the jth animal in the ith group, θ_i is the mean for group i, $\sigma^2 = 16$ is the variance, and the product is taken over all groups (i) and all observations within each group (j).

Now,to generate 20 sample observations from the joint posterior density using the Metropolis-Hastings algorithm:

Step 1: Define Likelihood and Prior: Given,

- Likelihood: $Y_{ij} \sim N(\theta_i, \sigma^2$
- Prior : N(62,4)

The likelihood function for each group can be written as:

$$\pi(data|\theta_1, \theta_2, \theta_3, \theta_4) = \prod_{i=1}^{4} \prod_{j=1}^{n_i} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(Y_{ij}-\theta_i)^2}{2\sigma^2}\right)$$

Step 2: Initialize Parameters and Variables:

- Set the number of groups,N (N=4)
- Initialize the number of samples,M (M=20)
- Initialize a matrix to store generated samples, $\theta_{samples}(dimension M \times N)$

Step 3: Metropolis-Hastings Algorithm:

For each sample m from 1 to M and for each group i from 1 to N

- **Initialize Current Value**: Set $\theta_i^{(m)}$ as the current value.
- **Calculate Log Likelihood and Log Prior** :
  $ln\pi(data|\theta_1, \theta_2, \theta_3, \theta_4) = \sum_{j=1}^{n_i} \log\left(\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(Y_{ij}-\theta_i^{(m)})^2}{2\sigma^2}\right)\right)$
  $ln\pi(\theta_i) = \sum_{i=1}^{N} \log\left(\frac{1}{\sqrt{2\pi \cdot 4}} \exp\left(-\frac{(\theta_i^{(m)}-62)^2}{2\cdot 4}\right)\right)$
- **Propose New Value:** Generate a proposed new value from a normal distribution: $\theta_i^* \sim \mathcal{N}(\theta_i^{(m)}, variance)$
- **Calculate Log Likelihood and Log Prior for Proposed Value:**

  Log Likelihood for Proposed Value: Similar to step 2 but using the proposed θ_i^*
  Log Prior for Proposed Value: Similar to step 2 but using the proposed θ_i^*
- **Calculate Acceptance Ratio:**
  Acceptance Ratio: =exp(proposed log likelihood+proposed log prior-log likelihood-log prior)
- **Accept or Reject**
  Generate a uniform random number U and if U $\leq$ acceptance ratio, set $\theta_i^{(m+1)} = \theta_i^{(m)}$ (reject the proposed value).
- Store the accepted or current value of $\theta_i^{(m+1)}$ in the matrix $\theta_{samples}$

Step 4: Output:

After running the loop for M samples, the matrix $\theta_{samples}$ contains M samples for each of the four groups' θ values, obtained using the Metropolis-Hastings algorithm. These samples approximate the joint posterior density of $(\theta_1, \theta_2, \theta_3, \theta_4)$ given the data and prior information.

The generated 20 samples for four different values of θs are as below:

```
         [,1]     [,2]     [,3]     [,4]
 [1,] 64.58765 59.37574 62.34521 57.40830
 [2,] 57.94120 62.39536 68.91752 64.72206
 [3,] 62.29111 56.09039 60.61085 62.79892
 [4,] 62.00520 61.90962 60.69318 62.89313
 [5,] 58.76205 65.21115 49.45290 67.56258
 [6,] 58.26250 62.34788 68.37163 61.47909
 [7,] 61.21470 63.61809 58.39525 64.78983
 [8,] 49.27971 59.94974 55.17690 59.73562
 [9,] 63.32462 58.80533 60.97746 57.19719
[10,] 64.06437 67.39013 57.40575 64.08362
[11,] 63.54245 60.57150 61.65884 68.13943
[12,] 65.03662 54.95713 67.89366 58.53373
[13,] 63.06807 62.96490 58.06497 59.22091
[14,] 59.86950 61.97735 61.97297 62.85959
[15,] 61.36697 57.16008 67.30220 54.89528
[16,] 63.41429 63.80290 62.48073 60.16419
[17,] 68.20537 68.54538 66.75312 59.41446
[18,] 64.59405 59.53116 59.53892 56.43223
[19,] 62.50081 60.54041 65.40522 56.56863
[20,] 60.05810 66.25001 63.40020 59.24621
```

```
 data <- list(
  groupA = c(62, 60, 63, 59),
  groupB = c(63, 67, 71, 64, 65, 66),
  groupC = c(68, 66, 71, 67, 67, 68, 68),
  groupD = c(56, 62, 60, 61, 63, 64, 63, 59)
)

num_groups <- length(data)

num_samples <- 20
theta_samples <- matrix(0, nrow = num_samples, ncol = num_groups)

sigma_squared <- 16
prior_mean <- 62
prior_sd <- 4

for (i in 1:num_samples) {
  for (group in 1:num_groups) {
    current_theta <- rnorm(1, prior_mean, prior_sd)

    log_likelihood <- sum(dnorm(data[[group]],
    mean = current_theta,
    sd = sqrt(sigma_squared), log = TRUE))

    log_prior <- sum(dnorm(current_theta, mean = prior_mean,
    sd = prior_sd, log = TRUE))

    proposed_theta <- rnorm(1, current_theta, 0.5)

    proposed_log_likelihood <- sum(dnorm(data[[group]],
    mean = proposed_theta,

    sd = sqrt(sigma_squared), log = TRUE))

    proposed_log_prior <- sum(dnorm(proposed_theta,
    mean = prior_mean,

    sd = prior_sd, log = TRUE))

    acceptance_ratio <- exp(proposed_log_likelihood +
    proposed_log_prior - log_likelihood - log_prior)

    if (runif(1) < acceptance_ratio) {
      current_theta <- proposed_theta
    }
```

```
    theta_samples[i, group] <- current_theta
  }
}
print(theta_samples)
```

3 Posterior Predictive Distribution:

A random sample of size 10 is drawn from a large population and the measurement on a certain characteristic is taken. Suppose the average measurement comes out to be 120. Assume that the characteristic is normally distributed with unknown mean θ and standard deviation 20. Suppose the prior distribution for *theta* is normal with mean 150 and standard deviation 40.

- Obtain the posterior distribution for θ
- A new unit is sampled at random from the same population and has a measure Y on the characteristic. Give a posterior predictive distribution for Y
- Derive a 95% posterior interval for θ and a 95% posterior predictive interval for Y.

Solution:

Mean of prior distribution (μ) = 150
Standard deviation of prior distribution (σ) = 40
And given:
Sample size (n) = 10
Sample mean = 120
Population standard deviation (σ) = 20

Let' s denote the following:

θ: true mean of the characteristic in the population.
μ: observed sample mean (120 in this case).
σ: known standard deviation of the characteristic (20 in this case).
μ_0 : mean of the prior distribution (150 in this case).
σ_0 : standard deviation of the prior distribution (40 in this case).
n: sample size (10 in this case).
Y: new measurement from the population.

a) Posterior Distribution for θ: The posterior distribution is given by Bayes' Theorem:

$$f(\theta|\text{data}) \propto f(\text{data}|\theta) \cdot f(\theta)$$

where f(θ|data) is the posterior distribution, f(data|θ) is the likelihood, and f(θ) is the prior distribution.

Given that the characteristic is normally distributed, the likelihood is:

$$f(\text{data}|\theta) \propto \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^{n}(x_i - \theta)^2\right)$$

The prior distribution is given as a normal distribution:

$$f(\text{data}|\theta) \propto \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^{n}(\theta - \mu_0)^2\right)$$

Multiplying the likelihood and the prior, we get the unnormalized posterior distribution:

$$f(\text{data}|\theta) \propto \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^{n}(x_i - \theta)^2\right) . \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^{n}(\theta - \mu_0)^2\right)$$

To simplify the expression and obtain a normal distribution, you can complete the square in the exponentials.

b) Posterior Predictive Distribution for Y given by:

f (Y|data) = ∫ f (Y|θ) · f(θ|data) dθ

where f (Y|θ) is the likelihood of the new measurement given θ. Since the characteristic is normally distributed, (Y|θ) is a normal distribution with mean θ and standard deviation σ.

The mean ($\mu_{posterior}$) and standard deviation ($\sigma_{posterior}$) of the posterior distribution can be calculated using the formulas for updating a normal distribution with new data. Here' s the form of the posterior distribution:
θ|data ~ N ($\mu_{posterior}$, ${\sigma_{posterior}}^2$)

The parameters of the posterior distribution are given by:

$$\mu_{posterior} = \frac{\frac{\mu_0}{\sigma_0^2} + \frac{\sum_{i=1}^{n} x_i}{\sigma^2}}{\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}}$$

$$\sigma_{posterior} = \sqrt{\frac{1}{\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}}}$$

In this context:
μ_0 : Mean of the prior distribution
σ_0 : Standard deviation of the prior distribution
n : Sample size
σ : Known standard deviation of the characteristic in the population
xi: Individual measurements from the sample

Output:

```
 > predictive_mean
[1] 126
> predictive_sd
[1] 26.83282
```

c) 95% Posterior Interval for θ and 95% Posterior Predictive Interval for Y: To obtain a credible interval for θ, we can find the quantiles of the posterior distribution. For Y, we can find the quantiles of the posterior predictive distribution.
Output:

```
> lower_posterior
[1] 90.9391
> upper_posterior
[1] 161.0609
> lower_predictive
[1] 73.40865
> upper_predictive
[1] 178.5914
```

code :

```
prior_mean <- 150
prior_sd <- 40
sample_mean <- 120
sample_sd <- 20
sample_size <- 10
# Calculate posterior distribution parameters
posterior_mean <- (prior_sd^2 * sample_mean + sample_sd^2 * prior_mean)
/ (prior_sd^2 + sample_sd^2)
posterior_sd <- sqrt((prior_sd^2 * sample_sd^2) /
(prior_sd^2 + sample_sd^2))
# Display posterior distribution parameters
posterior_mean
posterior_sd
# Calculate posterior predictive distribution parameters for Y
population_sd <- 20  # Given population standard deviation
predictive_mean <- posterior_mean
predictive_sd <- sqrt(posterior_sd^2 + population_sd^2)
# Display posterior predictive distribution parameters for Y
predictive_mean
predictive_sd
# Calculate 95% posterior interval for
lower_posterior <- qnorm(0.025, posterior_mean, posterior_sd)
upper_posterior <- qnorm(0.975, posterior_mean, posterior_sd)
# Calculate 95% posterior predictive interval for Y
lower_predictive <- qnorm(0.025, predictive_mean, predictive_sd)
upper_predictive <- qnorm(0.975, predictive_mean, predictive_sd)
# Display 95% posterior interval for  and 95% posterior predictive interval for Y
lower_posterior
upper_posterior
lower_predictive
upper_predictive
```